Detection in Magnetic Recording with Cross-Track Interference *

Mathew P. Vea José M.F. Moura

Data Storage Systems Center, Carnegie Mellon University

Abstract–We have analyzed the simple case of an isolated pulse with a single interfering pulse to get insight into the practial cross-track interference problem. We treat the problem in signal space to gain an intuitive geometric interpretation. We derive the Maximum Likelihood detector, the Linear Minimum Mean Squared Error (LMMSE) detector, and the minimum probability of error (mpe) linear detector. The mpe detector was found to closely approach the performance of the optimal ML nonlinear detector, and to significantly improve on the performance of the LMMSE detector as well a detector naively designed to match a channel without cross-track interference. These results provide insight and motivation to design equalizers to improve detector performance in the presence of such interference.

1 Introduction

Interference from information recorded off the track of interest is a significant factor limiting the performance of magnetic recording channels. This is especially true of hard disk storage systems, where the interference comes from adjacent tracks as well as "old information" in the guard band. This interference is not Gaussian, and so requires separate treatment from that of other noise sources in the magnetic recording channel. Cross-track or intertrack interference has been treated before, but often is assumed to have the same channel response as the on-track response[1]. In fact, the response of the channel to adjacent tracks will be different from its on-track response due to the side reading properties of the head [4]. This is due to a broader pulse response of the head when side reading.

The work reported here is an initial study of the cross-track interference problem. We study the case of an isolated pulse and a single interfering pulse with distinct pulse shapes. We have derived three detection schemes, the Maximum Likelihood detector and two linear schemes. These schemes are compared, at different interference levels and pulse shapes, to the performance of a detector which lacks knowledge of the interference.

*This material is based on work supported by the National Science Foundation under Grant No. ECD-8907068. The government has certain rights in this material

2 Problem Formulation

The isolated pulse problem involves the detection of a desired data symbol α from the observation $x(t)$, which includes the interfering symbol β and additive white Gaussian noise $n(t)$:

$$x(t) = \alpha p_1(t) + \beta p_2(t) + n(t). \quad (1)$$

The noise is white with spectral density σ^2. The desired symbol α takes on values $\{-1, 1\}$ with equal probability, β takes on equally likely values $\{-\rho, \rho\}$ and α and β are statistically independent. Since α and β are binary, the pulses $p_1(t)$ and $p_2(t)$ represent the pulse, or dibit, response of the magnetic recording channel. These pulses are characterized by their inner product:

$$\gamma = < p_1(t), p_2(t) > . \quad (2)$$

In this problem all symbols are real-valued, so the inner product is defined as:

$$< x(t), y(t) > = \int_{-\infty}^{\infty} x(t)y(t)dt.$$

The pulses are of unit energy, i.e. $< p_i(t), p_i(t) >= 1$. The isolated pulse problem as stated is completely characterized by three parameters: interference correlation γ, interference amplitude ρ, and noise power σ^2.

The continuous time problem of equation (1) can be transformed to a two-dimensional detection problem by passing the observed signal $x(t)$ through sampled filters which are matched to the pulse shapes $p_1(t)$ and $p_2(t)$. This filtering operation is equivalent to taking the inner product of $x(t)$ with $p_1(t)$ and $p_2(t)$. The familiar Gram-Schmidt procedure can be used to create a vector of orthonormal basis functions $\Phi(t)$ from $p_1(t)$ and $p_2(t)$:

$$\Phi(t) = \begin{bmatrix} \phi_1(t) \\ \phi_2(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{-\gamma}{\sqrt{1-\gamma^2}} & \frac{1}{\sqrt{1-\gamma^2}} \end{bmatrix} \begin{bmatrix} p_1(t) \\ p_2(t) \end{bmatrix}$$

$$= \mathbf{A}^{-1}\mathbf{P(t)}. \quad (3)$$

$\mathbf{P}(t)$ is the vector of pulse functions and $\mathbf{A}^{-1}$ is the orthonormalizing transformation such that $\mathbf{P}(t) = \mathbf{A}\Phi(t)$. The statistic $\mathbf{x}$ is formed as the result of the inner product of $x(t)$ and $\Phi(t)$:

$$\mathbf{x} =< \Phi(t), x(t) > . \quad (4)$$

The notation here is somewhat unusual: each element of $\mathbf{x}$ is equal to the inner product of the corresponding element time function of $\Phi(t)$ with $x(t)$.

The constituents of $\mathbf{x}$, x_1 and x_2, together form a sufficient statistic for the detection of α. This is easily shown to be true by contructing a complete set of orthonormal basis functions [3, p. 171] $\{\phi_i(t)\}$ whose first two members are $\phi_1(t)$ and $\phi_2(t)$. The statistics formed by taking the inner product with the higher order basis functions ($x_j =< \phi_j(t), x(t) >$ for $j > 2$) are independent of α, and therefore give no additional information for the detection of α. Thus, the continuous time observation $x(t)$ has been reduced to two dimensions without loss of information. The resulting statistic $\mathbf{x}$ is a function of the symbol vector $\Theta = [\alpha \ \beta]^T$ and the noise vector $\mathbf{n}$: [1]

$$\begin{aligned} \mathbf{x} &= < \Phi(t), \mathbf{P}^T(t)\Theta + n(t) > \\ &= < \Phi(t), \Phi^T(t)\mathbf{A}^T\Theta > + < \Phi(t), n(t) > \\ &= \mathbf{A}^T\Theta + \mathbf{n}. \end{aligned} \quad (5)$$

The noise $\mathbf{n}$ is zero-mean and Gaussian with variance $\sigma^2\mathbf{I}$. The two noise components n_1 and n_2 are independent as a result of the choice of an orthonormal basis.

It is evident from equation (5) that the probability distribution of $\mathbf{x}$ conditioned on Θ, $p(\mathbf{x}|\Theta)$, is Gaussian with mean $\mathbf{A}^T\Theta$ and variance $\sigma^2 I$. The detection problem can be visualized in signal space, as shown in Figure 1, with the conditional mean of $\mathbf{x}$ plotted for all values of α and β. The four different symbol pairs $\{\Theta_i\}$ shown in Figure 1 are defined to take on the following values of α and β:

$$\theta_1 = \begin{bmatrix} 1 \\ \rho \end{bmatrix}, \theta_2 = \begin{bmatrix} 1 \\ -\rho \end{bmatrix}, \theta_3 = \begin{bmatrix} -1 \\ \rho \end{bmatrix}, \theta_4 = \begin{bmatrix} -1 \\ -\rho \end{bmatrix} \quad (6)$$

The problem of detecting α is reduced to drawing a boundary in signal space to distinguish the region where $\alpha = 1$ is decided from the decision region for $\alpha = -1$. The detection schemes derived in the following section differ only in the criterion for which they draw this boundary.

[1] Again, the inner product here is applied to each element time function in the vector of time functions. Thus, $< \Phi(t), \Phi(t)^T(t) >= I$, the identity matrix

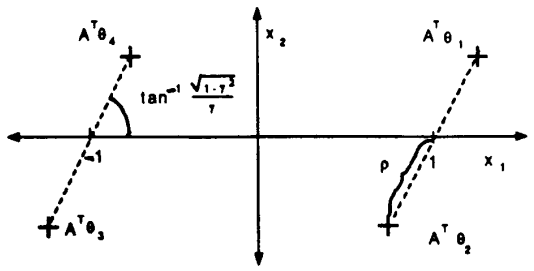

Figure 1: Signal space representation of isolated pulse detection

3 Detector design

The ideal detector has the minimum probability of error among all detection schemes. For the case of equally likely symbols, this detector is the Maximum Likelihood (ML) detector. The ML detector provides an upper bound on detector performance.

The decision boundary for the ML detector is drawn by comparing the likelihood ratio $\Lambda(\mathbf{x})$ to zero:

$$\Lambda(\mathbf{x}) = \frac{p(\mathbf{x}|\alpha = 1)}{p(\mathbf{x}|\alpha = -1)} \underset{<}{\overset{>}{}} 1. \quad (7)$$

Since this is a binary detection problem, there is only one decision threshold and the problem can be reduced to a single dimension. This problem is a composite problem because the interfering symbol β is an unwanted parameter. In composite problems, the effect of the unwanted parameter is averaged out by summing over all values of the parameter. In this case, there are two equiprobable values for β. The likelihood function is the ratio of two sums of two Gaussian functions:

$$\Lambda(\mathbf{x}) = \frac{p(\mathbf{x}|\Theta_1) + p(\mathbf{x}|\Theta_2)}{p(\mathbf{x}|\Theta_3) + p(\mathbf{x}|\Theta_4)}. \quad (8)$$

The likelihood ratio is clearly a nonlinear function of $\mathbf{x}$. The ML detector is impractical for this reason, but its probability of error is important to bound the performance of other detection schemes. The probability of error can be computed numerically by integrating the likelihood ratio over the decision region in the two-dimensional signal space:

$$P(e) = \int_{\Lambda(\mathbf{x})>0} p(\mathbf{x}|\alpha = -1)d\mathbf{x}. \quad (9)$$

3.1 Linear detectors

Although the signal space of Figure 1 is two dimensional, only one dimension is required to detect α. If

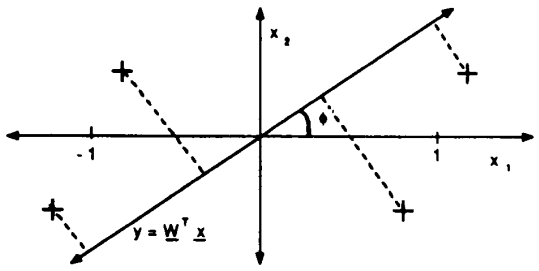

Figure 2: Linear detector signal space

the resulting scalar signal is a *linear* function of the statistic $\mathbf{x}$, then the detection scheme is linear. Specifically, the statistic $\mathbf{x}$ is reduced to the statistic y by the weight vector w:

$$y = \mathbf{w}^T\mathbf{x} \lessgtr T \tag{10}$$

and then y is compared to a threshold T to decide α. The decision boundary of a linear detector is the straight line whose equation is $y = \mathbf{w}^T\mathbf{x} = T$. Another interpretation of the linear detector is that it projects each point in the two-dimensional signal space onto a line perpendicular to the decision threshold, as shown in Figure 2.

A major advantage of linear detectors is that they require just one inner product, while nonlinear detectors require two. This can be shown by interchanging the order of the weighting and the inner product operators:

$$\begin{aligned} y = \mathbf{w}^T\mathbf{x} &= \mathbf{w}^T < \mathbf{\Phi}(t), x(t) > \\ &= < \mathbf{w}^T\mathbf{\Phi}(t), x(t) > . \end{aligned} \tag{11}$$

Thus two filters with impulse responses $\phi_1(t)$ and $\phi_2(t)$ can be replaced by a single filter whose impulse response is a weighted function of the matched filters: $\mathbf{w}^T\mathbf{\Phi}(t)$.

The probability of error of a linear detector is a function of the weight vector $\mathbf{w}$. The derivation is straightforward and relies on the fact that y is a linear function of x, so y given symbols $\mathbf{\Theta}$ is Gaussian. Specifically, $y|\mathbf{\Theta}$ has the distribution $N(\mathbf{w}^TA^T\mathbf{\Theta}, \sigma^2\mathbf{w}^T\mathbf{w})$. The problem was structured such that there are two symmetric symbol pairs: $\Theta_1 = -\Theta_4$ and $\Theta_2 = -\Theta_3$. The signal space for y is just a line, with the four conditional means representing the four possible symbol pairs. For the linear detector, the conditional means $\{\mathbf{w}^TA^T\Theta_i\}$ are correspondingly grouped into symmetric pairs, as shown in Figure 2. For the purpose of this derivation, we assume that $\mathbf{w}^TA^T\Theta_1 > 0$ and $\mathbf{w}^TA^T\Theta_2 > 0$. [2] In this case the optimum (ML) decision threshold will be zero. This threshold can be derived by constructing the likelihood ratio and noting that $p(y|\alpha = 1) = p(-y|\alpha = -1)$. The probability of error for a linear detector is therefore a simple function of the conditional means. Formally, the probability of error is equal to:

$$\begin{aligned} P(e) &= \frac{1}{2}P(e|\Theta_1) + \frac{1}{2}P(e|\Theta_2) \\ &= \frac{1}{2}Q(\frac{\mathbf{w}^TA^T\Theta_1}{\sigma\sqrt{\mathbf{w}^T\mathbf{w}}}) + \frac{1}{2}Q(\frac{\mathbf{w}^TA^T\Theta_2}{\sigma\sqrt{\mathbf{w}^T\mathbf{w}}}), \end{aligned} \tag{12}$$

where $Q(x)$ is an error function:

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-y^2/2}dy. \tag{13}$$

Equation (12) follows from the assumption of equiprobable symbols and the observation that $P(e|\Theta_1) = P(e|\Theta_4)$ and $P(e|\Theta_2) = P(e|\Theta_3)$ due to symmetry.

3.2 LMMSE detector

The derivation of the LMMSE weight vector follows from the Orthogonality Principle, which states that the error signal is orthogonal to the observations for the optimal $\mathbf{w}$. For the linear detector of equation (11), y functions as the estimate of α, and therefore the error signal is equal to $y - \alpha$. The Orthogonality Principle states that:

$$\begin{aligned} E\{(y-\alpha)\mathbf{x}\} &= 0 \\ \text{or} \qquad E\{y\mathbf{x}\} &= E\{\alpha\mathbf{x}\} \end{aligned} \tag{14}$$

The optimal weight vector $\mathbf{w}$ is derived by substituting the definitions for y (11) and $\mathbf{x}$ (5) into equation (14):

$$(A^TE\{\mathbf{\Theta}\mathbf{\Theta}^T\} + \sigma^2 I)\mathbf{w} = A^TE\{\alpha\mathbf{\Theta}\} \tag{15}$$

This equation follows from the assumptions that all symbols and noise are independent with zero mean. The final expression for $\mathbf{w}$ is found by substituting the statistics for α and β and solving the linear equations:

$$\mathbf{w} = \frac{1}{\sigma^4 + (1+\rho^2)\sigma^2 + \rho^2(1-\gamma^2)} \begin{bmatrix} \sigma^2 + (1+\gamma^2)\rho^2 \\ -\gamma\rho^2\sqrt{1-\gamma^2} \end{bmatrix} \tag{16}$$

[2] This assumption does not contrain our solution unduly. If one of $\mathbf{w}^TA^T\Theta_1$ and $\mathbf{w}^TA^T\Theta_2$ is negative and the other positive, the detector will have poor performance; this case is not of interest. If they are both negative for a given $\mathbf{w}$, the detector performance is identical to the case with weight vector $-\mathbf{w}$ where they are both positive.

3.3 Mpe linear detector

The LMMSE detector minimizes the probability of error if the only interference is Gaussian noise. It does not minimize probability of error for this case, because the interfering symbol β is not Gaussian. The linear detector which minimizes probability of error can be derived by making a few simple observations about equation (12). The probability of error is a sum of two error functions. For most cases, $P(e)$ will be dominated by the error function with the smaller (less negative) argument. Therefore a simple approximation to the minimum probability of error receiver is to maximize the minimum amplitude of the two arguments. That is, the weight vector $\mathbf{w}$ is chosen according to:

$$\max_{\mathbf{w}}\{\min(\mathbf{w}^T A^T \Theta_1, \mathbf{w}^T A^T \Theta_2)\}. \qquad (17)$$

A second observation about the arguments is that they are both of the form $\frac{\mathbf{w}^T A^T \Theta_i}{\sigma\sqrt{\mathbf{w}^T\mathbf{w}}}$, which is a simply a projection of $A^T\Theta_i$ onto the line $\mathbf{w}^T\mathbf{x} = 0$. These arguments are therefore independent of the magnitude of $\mathbf{w}$, and so also is the probability of error. Therefore we can assume that the magnitude of $\mathbf{w}$ is one and that it takes on the form:

$$\mathbf{w} = \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix} \qquad (18)$$

where ϕ is the angle which the projection line makes with the x-axis. Therefore the optimization problem is reduced to one dimension: choose ϕ to maximize equation (17). In fact, an exact value for ϕ can be computed by taking the derivative of the probability of error equation 12 with respect to ϕ, and setting it equal to zero. Equation (17) provides a more intuitive solution, which we will follow.

We have restricted ourselves to the case where both error function arguments are positive. When the two arguments are plotted as functions of ϕ, as shown in Figures 3 and 4, it is apparent that there are two regions of the plot:

$$\begin{aligned}
\text{Region I}: \quad & \frac{-(1+\rho\gamma)}{\rho\sqrt{1-\gamma^2}} < \tan\phi < \frac{-\gamma}{\sqrt{1-\gamma^2}} \\
\text{where} \quad & \mathbf{w}^T A^T \Theta_2 > \mathbf{w}^T A^T \Theta_1 > 0 \\
\text{Region II}: \quad & \frac{-\gamma}{\sqrt{1-\gamma^2}} < \tan\phi < \frac{1-\rho\gamma}{\rho\sqrt{1-\gamma^2}} \\
\text{where} \quad & \mathbf{w}^T A^T \Theta_1 > \mathbf{w}^T A^T \Theta_2 > 0 \qquad (19)
\end{aligned}$$

The value of ϕ that solves equation (17) will either occur at the intersection of the two argument functions, or at the maximum of one of the two functions. Since

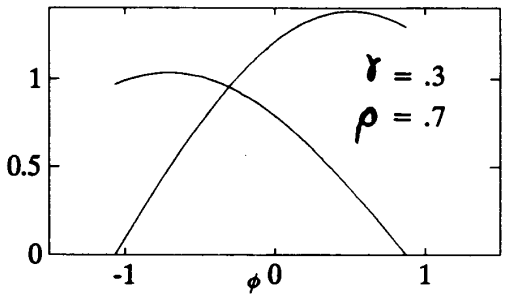

Figure 3: Case 1: distance metrics vs. ϕ

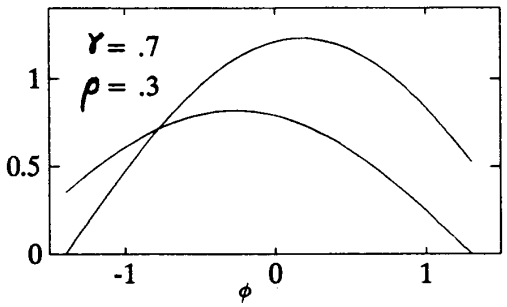

Figure 4: Case 2 distance metrics vs. ϕ

the arguments are projection funcitions, each is maximized when $\mathbf{w}$ is aligned with $A^T\Theta_i$. A little bit of analysis shows that the maximum of $\mathbf{w}^T A^T \Theta_1$ always occurs in region II, and therefore is never the minimum of the two arguments. The maximum of the second argument, $\mathbf{w}^T A^T \Theta_2$, will occur in region II if $\rho < \gamma$. So the design of the minimum probability linear detector follows one of two cases. In case 1, the optimal weight vector maximizes $\mathbf{w}^T A^T \Theta_2$. In case 2, the optimal weight vector occurs where the two error functions are equal. Specifically,:

$$\begin{aligned}
\rho > \gamma\text{: } \mathbf{w} &= \begin{bmatrix} \sqrt{1-\gamma^2} \\ -\gamma \end{bmatrix} \\
\rho < \gamma\text{: } \mathbf{w} &= \frac{1}{\sqrt{1-2\rho\gamma+\rho^2}} \begin{bmatrix} 1-\rho\gamma \\ \sqrt{1-\gamma^2} \end{bmatrix} \qquad (20)
\end{aligned}$$

The distance metrics from equation (12) are plotted for these two cases in Figures 3 and 4.

4 Performance Comparison

The probability of error for each of the three detection schemes, the ML, the LMMSE, and the minimum probability of error (mpe) detectors, were computed

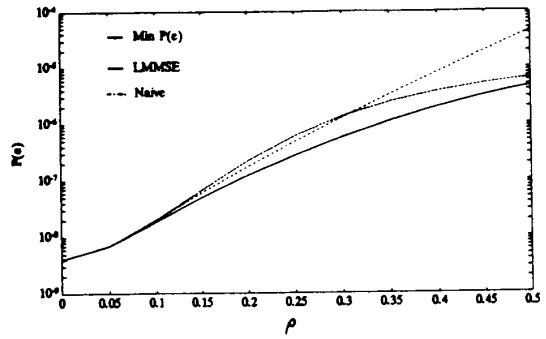

Figure 5: Detector performance vs. ρ

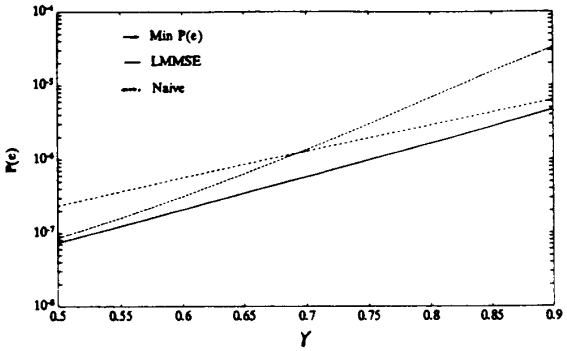

Figure 6: Detector performance vs. γ

for varying values of γ and ρ. In addition, the performance of a fourth detector was computed for comparison. This detector, referred to in figures as the "naive" detector, consists of an input filter with response $p_1(t)$ and an optimal slicer. Thus, it lacks knowledge of the cross-track response.

The performance of the mpe detector as plotted on the scale of Figures 5 and 6 is indistinguishable from that of the ML detector, so the ML detector was not included in the plots. This signifies that a linear detector is sufficient to achieve near-optimal performance for cases of interest to magnetic recording. Only when correlation γ approaches zero does the mpe performance diverge from the ML detector. The performance of the three schemes is plotted as a function of the interference correlation coefficient ρ in Figure 5, where $\gamma = .7$ and $\sigma^2 = .03$. Note the surprisingly poor performance of the LMMSE detector, which is outperformed by the naive detector at moderate levels of interference ($\rho < .3$). Figure 6 compares the detector performance as a function of γ, with $\rho = .3$. Again, note the poor performance of the LMMSE detector, which is outperformed by the naive detector at high interference correlation ($\gamma > .7$).

5 Summary

This study is a simplification of the cross-track interference problem. There may be more than one interfering track, for example when the head reads information form both the guard band and an adjacent track. The level of interference and its pulse shape will be neither known nor constant. Also, we have ignored intersymbol interference to focus on intertrack interference. Nonetheless, in reducing the problem to its essentials we have gained insight. We demonstrated that the LMMSE equalizer is suboptimal in the presence of intertrack interference. This is significant, since the MMSE criterion is commonly used to design adaptive filters and equalizers[2]. We also showed that significant improvement can result from incorporating knowledge of the cross-track response into the equalizer design. And finally, we characterized the interference problem concisely in terms of pulse shape and amplitude and gained an intuitive geometric description of intertrack interference in signal space.

References

[1] W.L. Abbott, J.M. Cioffi, and H.M. Thapar, "Performance of digital magnetic recording with equalization and offtrack interference", *IEEE Trans. on Magnetics*, vol. MAG-27, No.1, pp. 705-716, Jan. 1991.

[2] D.G. Messerschmitt, L.C. Barbosa, and T.D. Howell, "A study of sampling detectors for magnetic recording", Tech. Report RJ 4081, IBM Almaden Research Center, San Jose, Ca., July 1984.

[3] H.L. Van Trees, *Detection, Estimation, and Modulation Theory*, Wiley, 1968.

[4] M.P. Vea and J.M.F. Moura, "Magnetic recording channel model with intertrack interference", *IEEE Trans. on Magnetics*, vol. MAG-27, No. 6, Nov. 1991.